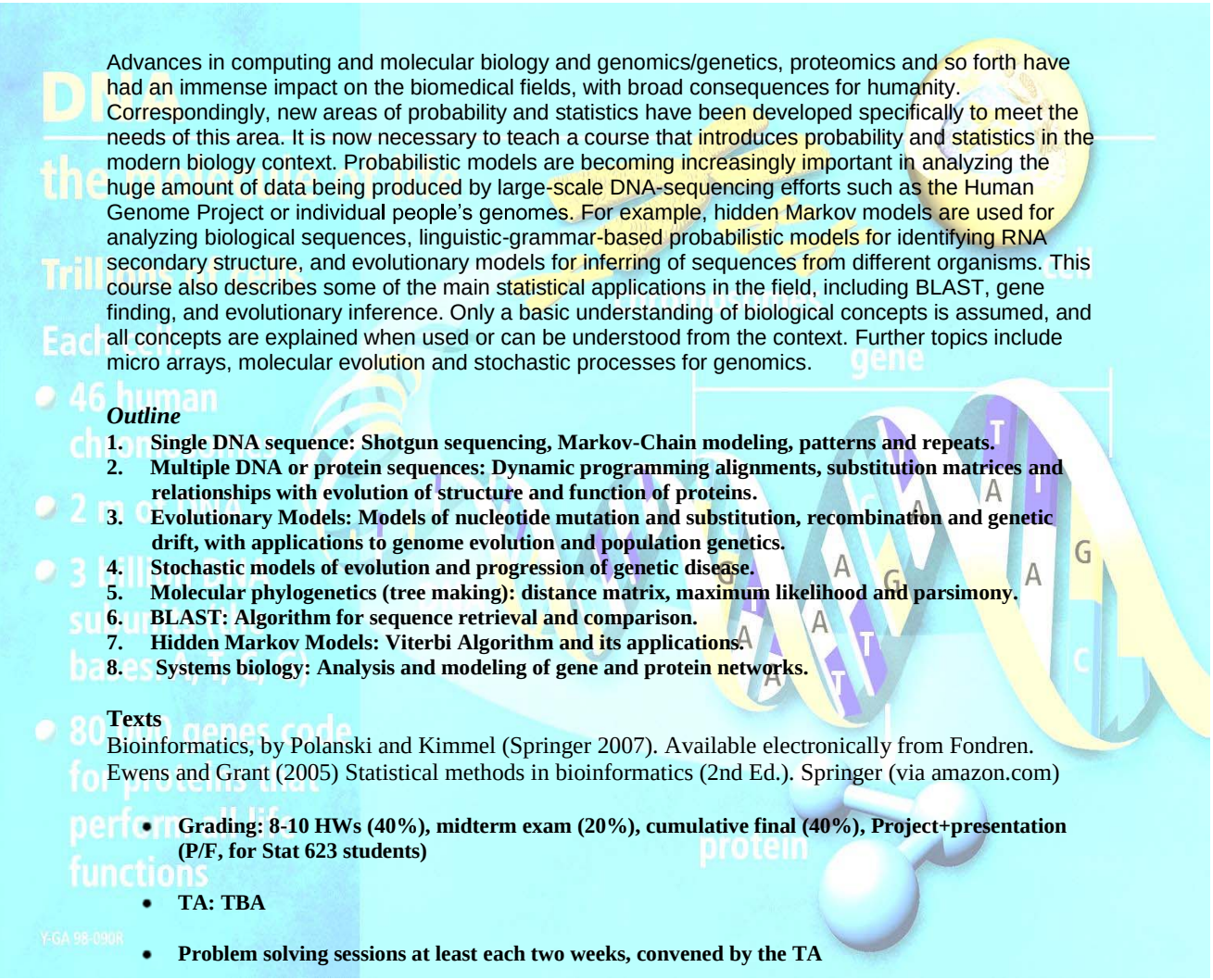
Course Announcement
STAT 423/623
Probability in Bioinformatics and Genetics
(and Systems Biology)
Spring 2015, TTr 10.50 – 12.05 AM
Room DH 1075



Advances in computing and molecular biology and genomics/genetics, proteomics and so forth have had an immense impact on the biomedical fields, with broad consequences for humanity. Correspondingly, new areas of probability and statistics have been developed specifically to meet the needs of this area. It is now necessary to teach a course that introduces probability and statistics in the modern biology context. Probabilistic models are becoming increasingly important in analyzing the huge amount of data being produced by large-scale DNA-sequencing efforts such as the Human Genome Project or individual people's genomes. For example, hidden Markov models are used for analyzing biological sequences, linguistic-grammar-based probabilistic models for identifying RNA secondary structure, and evolutionary models for inferring of sequences from different organisms. This course also describes some of the main statistical applications in the field, including BLAST, gene finding, and evolutionary inference. Only a basic understanding of biological concepts is assumed, and all concepts are explained when used or can be understood from the context. Further topics include micro arrays, molecular evolution and stochastic processes for genomics.

Outline

1. **Single DNA sequence: Shotgun sequencing, Markov-Chain modeling, patterns and repeats.**
2. **Multiple DNA or protein sequences: Dynamic programming alignments, substitution matrices and relationships with evolution of structure and function of proteins.**
3. **Evolutionary Models: Models of nucleotide mutation and substitution, recombination and genetic drift, with applications to genome evolution and population genetics.**
4. **Stochastic models of evolution and progression of genetic disease.**
5. **Molecular phylogenetics (tree making): distance matrix, maximum likelihood and parsimony.**
6. **BLAST: Algorithm for sequence retrieval and comparison.**
7. **Hidden Markov Models: Viterbi Algorithm and its applications.**
8. **Systems biology: Analysis and modeling of gene and protein networks.**

Texts

Bioinformatics, by Polanski and Kimmel (Springer 2007). Available electronically from Fondren.
Ewens and Grant (2005) Statistical methods in bioinformatics (2nd Ed.). Springer (via amazon.com)

- **Grading: 8-10 HWs (40%), midterm exam (20%), cumulative final (40%), Project+presentation (P/F, for Stat 623 students)**
- **TA: TBA**
- **Problem solving sessions at least each two weeks, convened by the TA**

For more information contact:
David Wallace-Bradley (dw19@rice.edu)